

eval

The main task, comprising all test items.

number of items in task:	8448
--------------------------	------

	all senses	main senses only
average polysemy:	10.372	7.207

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.916	1.512

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.965 (0.963)	0.968 (0.967)	0.970 (0.968)
best system	0.771 (0.771)	0.797 (0.797)	0.814 (0.813)
average of systems	0.550 (0.376)	0.632 (0.410)	0.661 (0.426)
worst system	0.205 (0.162)	0.315 (0.248)	0.338 (0.267)
best <u>baseline</u>	0.691 (0.689)	0.720 (0.719)	0.741 (0.739)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.559 (0.519)	0.616 (0.031)	0.650 (0.198)
average of systems	0.406 (0.222)	0.531 (0.264)	0.563 (0.280)
worst system	0.205 (0.162)	0.315 (0.248)	0.338 (0.267)

best <u>baseline</u>	0.550 (0.548)	0.584 (0.582)	0.600 (0.597)
----------------------	----------------------	----------------------	----------------------

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.771 (0.771)	0.797 (0.797)	0.814 (0.813)
average of systems	0.661 (0.540)	0.707 (0.570)	0.733 (0.588)
worst system	0.433 (0.138)	0.588 (0.188)	0.627 (0.200)
best <u>baseline</u>	0.691 (0.689)	0.720 (0.719)	0.741 (0.739)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.127)	0.733 (0.143)	0.767 (0.150)
average of systems	0.527 (0.091)	0.637 (0.109)	0.663 (0.113)
worst system	0.405 (0.056)	0.540 (0.075)	0.559 (0.077)

Detailed results**trainable**

The subset of test items in files of words for which corpus training data was supplied.

number of items in task:	7446
--------------------------	------

	all senses	main senses only
average polysemy:	10.788	7.432

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.962	1.551

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.963 (0.962)	0.967 (0.966)	0.969 (0.967)
best system	0.763 (0.756)	0.788 (0.780)	0.801 (0.793)
average of systems	0.548 (0.393)	0.622 (0.424)	0.655 (0.442)
worst system	0.199 (0.156)	0.306 (0.240)	0.333 (0.261)
best <u>baseline</u>	0.709 (0.708)	0.735 (0.734)	0.759 (0.758)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.574 (0.533)	0.617 (0.572)	0.648 (0.600)
average of systems	0.400 (0.219)	0.512 (0.256)	0.551 (0.275)
worst system	0.199 (0.156)	0.306 (0.240)	0.333 (0.261)
best <u>baseline</u>	0.549 (0.547)	0.582 (0.579)	0.599 (0.596)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.763 (0.756)	0.788 (0.780)	0.801 (0.793)
average of systems	0.666 (0.574)	0.705 (0.601)	0.733 (0.623)
worst system	0.438 (0.128)	0.564 (0.165)	0.612 (0.179)
best <u>baseline</u>	0.709 (0.708)	0.735 (0.734)	0.759 (0.758)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.144)	0.733 (0.162)	0.767 (0.170)
average of systems	0.506 (0.093)	0.622 (0.111)	0.651 (0.117)
worst system	0.363 (0.043)	0.510 (0.060)	0.535 (0.063)

Detailed results**untrainable**

The complement of trainable.

number of items in task:	1002
--------------------------	------

	all senses	main senses only
average polysemy:	7.276	5.533

	fine-grained	coarse-grained
average <u>entropy</u> :	1.568	1.220

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.973 (0.973)	0.974 (0.974)	0.975 (0.975)
best system	0.882 (0.882)	0.930 (0.930)	0.930 (0.930)
average of systems	0.512 (0.343)	0.660 (0.414)	0.660 (0.415)

worst system	0.248 (0.207)	0.375 (0.312)	0.376 (0.313)
best <u>baseline</u>	0.626 (0.626)	0.709 (0.709)	0.711 (0.711)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.559 (0.228)	0.742 (0.066)	0.742 (0.066)
average of systems	0.435 (0.243)	0.618 (0.320)	0.618 (0.321)
worst system	0.248 (0.207)	0.375 (0.312)	0.376 (0.313)
best <u>baseline</u>	0.626 (0.626)	0.709 (0.709)	0.711 (0.711)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.882 (0.882)	0.930 (0.930)	0.930 (0.930)
average of systems	0.608 (0.498)	0.717 (0.569)	0.719 (0.570)
worst system	0.415 (0.216)	0.590 (0.587)	0.592 (0.589)
best <u>baseline</u>	0.556 (0.553)	0.604 (0.602)	0.606 (0.603)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.533 (0.154)	0.631 (0.183)	0.631 (0.183)
average of systems	0.533 (0.154)	0.631 (0.183)	0.631 (0.183)
worst system	0.533 (0.154)	0.631 (0.183)	0.631 (0.183)

Detailed results

multi-word

The subset of test items tagged with sense tags for word forms that are not derivable from the root form of a test word by a regular morphological process.

number of items in task:	800
--------------------------	-----

	all senses	main senses only
average polysemy:	16.862	12.594

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.499	2.013

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.984 (0.982)	0.986 (0.984)	0.986 (0.984)
best system	0.897 (0.870)	0.914 (0.886)	0.921 (0.894)
average of systems	0.529 (0.399)	0.592 (0.432)	0.671 (0.471)
worst system	0.000 (0.000)	0.200 (0.002)	0.347 (0.291)
best <u>baseline</u>	0.815 (0.658)	0.832 (0.672)	0.861 (0.696)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.572 (0.309)	0.689 (0.372)	0.794 (0.429)
average of systems	0.289 (0.197)	0.381 (0.229)	0.515 (0.279)

worst system	0.000 (0.000)	0.200 (0.002)	0.347 (0.291)
best <u>baseline</u>	0.801 (0.800)	0.823 (0.822)	0.843 (0.842)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.897 (0.870)	0.914 (0.886)	0.921 (0.894)
average of systems	0.675 (0.583)	0.722 (0.620)	0.765 (0.655)
worst system	0.296 (0.141)	0.337 (0.160)	0.370 (0.176)
best <u>baseline</u>	0.815 (0.658)	0.832 (0.672)	0.861 (0.696)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.739 (0.238)	0.783 (0.252)	0.831 (0.268)
average of systems	0.735 (0.204)	0.765 (0.213)	0.808 (0.226)
worst system	0.731 (0.170)	0.747 (0.174)	0.785 (0.183)

Detailed results**simple-word**

The complement of multi-word.

number of items in task:	7648
--------------------------	------

	all senses	main senses only
average polysemy:	9.693	6.644

	<u>fine-grained</u>	<u>coarse-grained</u>
average entropy:	1.855	1.460

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.963 (0.961)	0.966 (0.965)	0.968 (0.966)
best system	0.761 (0.761)	0.787 (0.787)	0.804 (0.804)
average of systems	0.548 (0.374)	0.634 (0.408)	0.660 (0.422)
worst system	0.200 (0.156)	0.320 (0.251)	0.338 (0.265)
best <u>baseline</u>	0.678 (0.677)	0.709 (0.708)	0.730 (0.729)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.576 (0.544)	0.630 (0.595)	0.677 (0.194)
average of systems	0.411 (0.224)	0.541 (0.267)	0.569 (0.280)
worst system	0.200 (0.156)	0.320 (0.251)	0.338 (0.265)
best <u>baseline</u>	0.524 (0.522)	0.564 (0.564)	0.605 (0.605)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.761 (0.761)	0.787 (0.787)	0.804 (0.804)
average of systems	0.662 (0.536)	0.708 (0.564)	0.732 (0.581)
worst system	0.456 (0.138)	0.611 (0.538)	0.623 (0.548)

best <u>baseline</u>	0.678 (0.677)	0.709 (0.708)	0.730 (0.729)
----------------------	----------------------	----------------------	----------------------

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.633 (0.115)	0.724 (0.132)	0.756 (0.137)
average of systems	0.488 (0.080)	0.612 (0.098)	0.636 (0.101)
worst system	0.344 (0.044)	0.501 (0.064)	0.516 (0.066)

Detailed results**unassignable**

The subset of test items tagged with an UNASSIGNABLE sense tag in the key.

number of items in task:	35
--------------------------	----

	all senses	main senses only
average polysemy:	11.771	8.086

	fine-grained	coarse-grained
average entropy:	2.415	1.946

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.938 (0.857)	0.938 (0.857)	0.938 (0.857)
best system	0.400 (0.400)	0.509 (0.087)	0.509 (0.087)

average of systems	0.261 (0.166)	0.317 (0.190)	0.324 (0.196)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.343 (0.343)	0.371 (0.371)	0.371 (0.371)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.364 (0.138)	0.509 (0.087)	0.509 (0.087)
average of systems	0.230 (0.126)	0.296 (0.151)	0.299 (0.153)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.343 (0.343)	0.371 (0.371)	0.371 (0.371)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.400 (0.400)	0.500 (0.086)	0.500 (0.086)
average of systems	0.311 (0.222)	0.368 (0.251)	0.379 (0.260)
worst system	0.185 (0.143)	0.232 (0.186)	0.250 (0.200)
best <u>baseline</u>	0.200 (0.200)	0.250 (0.186)	0.269 (0.200)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.205 (0.010)	0.205 (0.010)	0.205 (0.010)
average of systems	0.102 (0.005)	0.102 (0.005)	0.102 (0.005)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

Detailed results**proper**

The subset of test items tagged with a PROPER NOUN sense tag in the key.

number of items in task:	286
--------------------------	-----

	all senses	main senses only
average polysemy:	12.014	8.899

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.035	1.671

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.991 (0.981)	0.993 (0.983)	0.993 (0.983)
best system	0.884 (0.881)	0.930 (0.927)	0.930 (0.927)
average of systems	0.510 (0.267)	0.606 (0.305)	0.613 (0.308)
worst system	0.026 (0.002)	0.263 (0.017)	0.263 (0.017)
best <u>baseline</u>	0.756 (0.325)	0.756 (0.325)	0.756 (0.325)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.516 (0.065)	0.720 (0.629)	0.724 (0.633)
average of systems	0.308 (0.140)	0.476 (0.203)	0.481 (0.206)

worst system	0.026 (0.002)	0.263 (0.017)	0.263 (0.017)
best <u>baseline</u>	0.556 (0.556)	0.633 (0.633)	0.643 (0.643)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.884 (0.881)	0.930 (0.927)	0.930 (0.927)
average of systems	0.682 (0.395)	0.733 (0.419)	0.741 (0.423)
worst system	0.382 (0.381)	0.399 (0.397)	0.407 (0.406)
best <u>baseline</u>	0.756 (0.325)	0.756 (0.325)	0.756 (0.325)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.481 (0.060)	0.501 (0.063)	0.525 (0.066)
average of systems	0.389 (0.067)	0.426 (0.075)	0.438 (0.076)
worst system	0.296 (0.073)	0.352 (0.087)	0.352 (0.087)

Detailed results**nouns**

All test items in files with **-n** suffix.

number of items in task:	2756
--------------------------	------

	all senses	main senses only
average polysemy:	9.167	5.381

	<u>fine-grained</u>	<u>coarse-grained</u>
average entropy:	1.740	1.167

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.973 (0.973)	0.975 (0.975)	0.977 (0.977)
best system	0.850 (0.850)	0.886 (0.886)	0.918 (0.918)
average of systems	0.594 (0.463)	0.702 (0.529)	0.749 (0.565)
worst system	0.214 (0.174)	0.384 (0.311)	0.437 (0.354)
best <u>baseline</u>	0.738 (0.569)	0.815 (0.629)	0.879 (0.679)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.637 (0.581)	0.773 (0.083)	0.811 (0.087)
average of systems	0.445 (0.298)	0.608 (0.384)	0.662 (0.422)
worst system	0.214 (0.174)	0.384 (0.311)	0.437 (0.354)
best <u>baseline</u>	0.628 (0.625)	0.675 (0.672)	0.731 (0.731)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.850 (0.850)	0.886 (0.886)	0.918 (0.918)
average of systems	0.712 (0.619)	0.772 (0.673)	0.817 (0.711)
worst system	0.433 (0.424)	0.588 (0.576)	0.627 (0.614)

<u>best baseline</u>	0.738 (0.569)	0.815 (0.629)	0.879 (0.679)
----------------------	----------------------	----------------------	----------------------

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.388)	0.733 (0.438)	0.767 (0.459)
average of systems	0.562 (0.264)	0.699 (0.316)	0.730 (0.331)
worst system	0.476 (0.139)	0.666 (0.194)	0.694 (0.202)

Detailed results**all-nouns**

All test items in nouns, plus test items in files with **-p** suffix that were tagged with noun sense tags in the key.

number of items in task:	3792
--------------------------	------

	all senses	main senses only
average polysemy:	10.949	7.445

	fine-grained	coarse-grained
average <u>entropy</u> :	1.832	1.363

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.976 (0.976)	0.978 (0.978)	0.980 (0.980)

best system	0.839 (0.839)	0.872 (0.872)	0.896 (0.896)
average of systems	0.582 (0.427)	0.671 (0.478)	0.708 (0.504)
worst system	0.235 (0.203)	0.359 (0.309)	0.395 (0.340)
best <u>baseline</u>	0.746 (0.558)	0.804 (0.602)	0.852 (0.638)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.624 (0.570)	0.689 (0.629)	0.727 (0.664)
average of systems	0.422 (0.267)	0.554 (0.334)	0.596 (0.362)
worst system	0.235 (0.203)	0.359 (0.309)	0.395 (0.340)
best <u>baseline</u>	0.564 (0.561)	0.604 (0.601)	0.642 (0.642)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.839 (0.839)	0.872 (0.872)	0.896 (0.896)
average of systems	0.712 (0.584)	0.764 (0.625)	0.799 (0.653)
worst system	0.433 (0.308)	0.588 (0.419)	0.627 (0.447)
best <u>baseline</u>	0.746 (0.558)	0.804 (0.602)	0.852 (0.638)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.282)	0.733 (0.318)	0.767 (0.333)
average of systems	0.528 (0.204)	0.638 (0.242)	0.664 (0.252)
worst system	0.408 (0.125)	0.543 (0.166)	0.562 (0.172)

Detailed results**verbs**

All test items in files with -v suffix.

number of items in task:	2501
--------------------------	------

	all senses	main senses only
average polysemy:	7.791	4.994

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.859	1.496

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.950 (0.947)	0.955 (0.952)	0.957 (0.954)
best system	0.705 (0.697)	0.741 (0.733)	0.755 (0.747)
average of systems	0.545 (0.500)	0.583 (0.532)	0.602 (0.549)
worst system	0.155 (0.106)	0.259 (0.178)	0.272 (0.187)
best <u>baseline</u>	0.701 (0.700)	0.727 (0.725)	0.746 (0.744)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.469 (0.443)	0.526 (0.497)	0.537 (0.507)
average of systems	0.390 (0.316)	0.448 (0.361)	0.465 (0.375)
worst system	0.155 (0.106)	0.259 (0.178)	0.272 (0.187)
best <u>baseline</u>	0.547 (0.545)	0.582 (0.579)	0.592 (0.589)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.705 (0.697)	0.741 (0.733)	0.755 (0.747)
average of systems	0.622 (0.592)	0.650 (0.618)	0.671 (0.637)
worst system	0.392 (0.392)	0.414 (0.414)	0.426 (0.425)
best <u>baseline</u>	0.701 (0.700)	0.727 (0.725)	0.746 (0.744)

Detailed results**all-verbs**

All test items in verbs, plus test items in files with **-p** suffix that were tagged with verb sense tags in the key.

number of items in task:	2907
--------------------------	------

	all senses	main senses only
average polysemy:	10.821	7.733

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.056	1.723

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.954 (0.951)	0.958 (0.956)	0.960 (0.957)
best system	0.714 (0.707)	0.748 (0.741)	0.761 (0.754)
average of systems	0.457 (0.419)	0.486 (0.444)	0.500 (0.458)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.676 (0.675)	0.699 (0.697)	0.717 (0.715)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.490 (0.460)	0.543 (0.509)	0.554 (0.519)
average of systems	0.278 (0.226)	0.316 (0.255)	0.327 (0.264)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.541 (0.539)	0.574 (0.572)	0.583 (0.581)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.714 (0.707)	0.748 (0.741)	0.761 (0.754)
average of systems	0.628 (0.597)	0.653 (0.621)	0.672 (0.639)
worst system	0.431 (0.430)	0.451 (0.450)	0.460 (0.460)
best <u>baseline</u>	0.676 (0.675)	0.699 (0.697)	0.717 (0.715)

Detailed results**adjectives**

All test items in files with **-a** suffix.

number of items in task:	1406
--------------------------	------

	all senses	main senses only
average polysemy:	6.760	4.576

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.658	1.236

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.966 (0.965)	0.972 (0.972)	0.973 (0.973)
best system	0.782 (0.756)	0.797 (0.771)	0.799 (0.772)
average of systems	0.614 (0.543)	0.652 (0.575)	0.667 (0.587)
worst system	0.181 (0.129)	0.312 (0.223)	0.323 (0.230)
best <u>baseline</u>	0.718 (0.717)	0.737 (0.735)	0.740 (0.738)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.557 (0.532)	0.590 (0.564)	0.629 (0.601)
average of systems	0.448 (0.368)	0.503 (0.411)	0.525 (0.429)
worst system	0.181 (0.129)	0.312 (0.223)	0.323 (0.230)
best <u>baseline</u>	0.681 (0.681)	0.694 (0.694)	0.709 (0.709)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.782 (0.756)	0.797 (0.771)	0.799 (0.772)
average of systems	0.697 (0.631)	0.726 (0.657)	0.737 (0.666)
worst system	0.589 (0.538)	0.631 (0.576)	0.634 (0.579)
best <u>baseline</u>	0.718 (0.717)	0.737 (0.735)	0.740 (0.738)

Detailed results**all-adjectives**

All test items in adjectives, plus test items in files with **-p** suffix that were tagged with adjective sense tags in the key.

number of items in task:	1750
--------------------------	------

	all senses	main senses only
average polysemy:	8.430	5.868

	<u>fine-grained</u>	<u>coarse-grained</u>
average entropy:	1.867	1.490

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.957 (0.956)	0.962 (0.962)	0.963 (0.963)
best system	0.755 (0.734)	0.766 (0.746)	0.768 (0.747)
average of systems	0.599 (0.533)	0.631 (0.561)	0.644 (0.571)
worst system	0.214 (0.163)	0.320 (0.244)	0.329 (0.251)
best <u>baseline</u>	0.688 (0.686)	0.703 (0.701)	0.705 (0.704)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.536 (0.505)	0.563 (0.530)	0.595 (0.560)
average of systems	0.438 (0.357)	0.486 (0.395)	0.504 (0.410)
worst system	0.214 (0.163)	0.320 (0.244)	0.329 (0.251)
best <u>baseline</u>	0.615 (0.615)	0.627 (0.627)	0.629 (0.629)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.755 (0.734)	0.766 (0.746)	0.768 (0.747)
average of systems	0.680 (0.621)	0.704 (0.643)	0.713 (0.652)
worst system	0.606 (0.563)	0.639 (0.594)	0.642 (0.597)
best <u>baseline</u>	0.688 (0.686)	0.703 (0.701)	0.705 (0.704)

Detailed results**indeterminates**

All test items in files with **-p** suffix; the part of speech of the word to be disambiguated has not been predetermined.

number of items in task:	1785
--------------------------	------

	all senses	main senses only
average polysemy:	18.692	15.199

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.468	2.284

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.970 (0.969)	0.972 (0.970)	0.973 (0.971)
best system	0.775 (0.775)	0.793 (0.793)	0.797 (0.797)
average of systems	0.536 (0.423)	0.551 (0.435)	0.553 (0.437)
worst system	0.249 (0.051)	0.260 (0.019)	0.260 (0.019)
best <u>baseline</u>	0.656 (0.531)	0.658 (0.533)	0.661 (0.535)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.572 (0.518)	0.633 (0.573)	0.635 (0.575)
average of systems	0.352 (0.231)	0.373 (0.249)	0.375 (0.251)
worst system	0.257 (0.246)	0.260 (0.019)	0.260 (0.019)
best <u>baseline</u>	0.425 (0.424)	0.444 (0.443)	0.445 (0.444)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.775 (0.775)	0.793 (0.793)	0.797 (0.797)
average of systems	0.694 (0.594)	0.705 (0.604)	0.707 (0.606)
worst system	0.555 (0.555)	0.567 (0.567)	0.569 (0.569)
best <u>baseline</u>	0.656 (0.531)	0.658 (0.533)	0.661 (0.535)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.249 (0.051)	0.262 (0.054)	0.262 (0.054)
average of systems	0.249 (0.051)	0.262 (0.054)	0.262 (0.054)
worst system	0.249 (0.051)	0.262 (0.054)	0.262 (0.054)

Detailed results**determinates**

All test items for ambiguous words with a predetermined part of speech; the complement of

indeterminates.

number of items in task:	6663
--------------------------	------

	all senses	main senses only
average polysemy:	8.143	5.066

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.768	1.305

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.963 (0.962)	0.967 (0.966)	0.969 (0.967)
best system	0.770 (0.769)	0.802 (0.793)	0.818 (0.818)
average of systems	0.558 (0.388)	0.656 (0.428)	0.690 (0.449)
worst system	0.187 (0.139)	0.326 (0.242)	0.357 (0.265)
best <u>baseline</u>	0.720 (0.719)	0.753 (0.752)	0.779 (0.778)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.556 (0.519)	0.773 (0.034)	0.811 (0.036)
average of systems	0.421 (0.232)	0.573 (0.282)	0.612 (0.303)
worst system	0.187 (0.139)	0.326 (0.242)	0.357 (0.265)
best <u>baseline</u>	0.584 (0.581)	0.622 (0.619)	0.641 (0.638)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.770 (0.769)	0.802 (0.793)	0.818 (0.818)
average of systems	0.661 (0.552)	0.711 (0.587)	0.741 (0.610)
worst system	0.433 (0.176)	0.588 (0.238)	0.626 (0.562)
best <u>baseline</u>	0.720 (0.719)	0.753 (0.752)	0.779 (0.778)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.160)	0.733 (0.181)	0.767 (0.190)
average of systems	0.562 (0.108)	0.699 (0.130)	0.730 (0.137)
worst system	0.476 (0.057)	0.666 (0.080)	0.694 (0.084)

Detailed results**trainable-nouns**

The intersection of nouns and trainable.

number of items in task:	2199
--------------------------	------

	all senses	main senses only
average polysemy:	10.067	5.675

	fine-grained	coarse-grained
average <u>entropy</u> :	1.892	1.271

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.976 (0.975)	0.978 (0.978)	0.980 (0.980)
best system	0.833 (0.833)	0.869 (0.869)	0.909 (0.909)
average of systems	0.594 (0.501)	0.692 (0.565)	0.748 (0.609)
worst system	0.204 (0.160)	0.373 (0.292)	0.440 (0.345)
best <u>baseline</u>	0.751 (0.751)	0.815 (0.788)	0.879 (0.850)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.642 (0.583)	0.753 (0.084)	0.800 (0.089)
average of systems	0.448 (0.298)	0.594 (0.377)	0.661 (0.425)
worst system	0.204 (0.160)	0.373 (0.292)	0.440 (0.345)
best <u>baseline</u>	0.615 (0.612)	0.659 (0.656)	0.706 (0.706)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.833 (0.833)	0.869 (0.869)	0.909 (0.909)
average of systems	0.713 (0.687)	0.767 (0.739)	0.819 (0.785)
worst system	0.438 (0.434)	0.564 (0.559)	0.612 (0.606)
best <u>baseline</u>	0.751 (0.751)	0.815 (0.788)	0.879 (0.850)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.486)	0.733 (0.549)	0.767 (0.575)
average of systems	0.535 (0.304)	0.678 (0.365)	0.712 (0.383)
worst system	0.420 (0.122)	0.622 (0.181)	0.657 (0.191)

Detailed results**all-trainable-nouns**

The intersection of all-nouns and trainable.

number of items in task:	2914
--------------------------	------

	<u>all senses</u>	<u>main senses only</u>
average polysemy:	11.963	8.000

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.898	1.406

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.978 (0.977)	0.980 (0.980)	0.982 (0.981)
best system	0.829 (0.829)	0.856 (0.856)	0.887 (0.887)
average of systems	0.588 (0.480)	0.668 (0.529)	0.713 (0.562)
worst system	0.231 (0.193)	0.353 (0.294)	0.401 (0.335)

best baseline	0.754 (0.754)	0.804 (0.783)	0.852 (0.830)
---------------	----------------------	----------------------	----------------------

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.687 (0.623)	0.734 (0.666)	0.783 (0.711)
average of systems	0.426 (0.276)	0.544 (0.337)	0.597 (0.373)
worst system	0.231 (0.193)	0.353 (0.294)	0.401 (0.335)
best baseline	0.571 (0.568)	0.607 (0.604)	0.639 (0.636)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.829 (0.829)	0.856 (0.856)	0.887 (0.887)
average of systems	0.723 (0.672)	0.769 (0.713)	0.810 (0.748)
worst system	0.438 (0.327)	0.564 (0.422)	0.612 (0.457)
best baseline	0.754 (0.754)	0.804 (0.783)	0.852 (0.830)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.367)	0.733 (0.414)	0.767 (0.434)
average of systems	0.507 (0.238)	0.623 (0.284)	0.653 (0.297)
worst system	0.366 (0.110)	0.513 (0.154)	0.539 (0.161)

Detailed results

untrainable-nouns

The intersection of nouns and untrainable.

number of items in task:	557
--------------------------	-----

	all senses	main senses only
average polysemy:	5.616	4.219

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.140	0.755

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.964 (0.964)	0.965 (0.965)	0.966 (0.966)
best system	0.916 (0.916)	0.954 (0.954)	0.955 (0.955)
average of systems	0.543 (0.419)	0.725 (0.524)	0.727 (0.526)
worst system	0.250 (0.230)	0.422 (0.387)	0.425 (0.390)
best <u>baseline</u>	0.756 (0.756)	0.828 (0.828)	0.831 (0.831)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.617 (0.569)	0.865 (0.081)	0.865 (0.081)
average of systems	0.428 (0.302)	0.665 (0.410)	0.666 (0.411)
worst system	0.250 (0.230)	0.422 (0.387)	0.425 (0.390)

best <u>baseline</u>	0.756 (0.756)	0.828 (0.828)	0.831 (0.831)
----------------------	----------------------	----------------------	----------------------

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.916 (0.916)	0.954 (0.954)	0.955 (0.955)
average of systems	0.668 (0.601)	0.787 (0.711)	0.789 (0.713)
worst system	0.415 (0.388)	0.691 (0.645)	0.692 (0.646)
best <u>baseline</u>	0.681 (0.680)	0.743 (0.741)	0.746 (0.744)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.691 (0.205)	0.836 (0.248)	0.836 (0.248)
average of systems	0.691 (0.205)	0.836 (0.248)	0.836 (0.248)
worst system	0.691 (0.205)	0.836 (0.248)	0.836 (0.248)

Detailed results**all-untrainable-nouns**

The intersection of all-nouns and untrainable.

number of items in task:	878
--------------------------	-----

	<u>all senses</u>	<u>main senses only</u>
average polysemy:	7.584	5.601

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.614	1.217

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.972 (0.972)	0.973 (0.973)	0.973 (0.973)
best system	0.871 (0.871)	0.925 (0.925)	0.926 (0.926)
average of systems	0.496 (0.340)	0.646 (0.419)	0.647 (0.420)
worst system	0.249 (0.236)	0.376 (0.356)	0.377 (0.358)
best <u>baseline</u>	0.603 (0.603)	0.697 (0.697)	0.699 (0.699)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.498 (0.326)	0.742 (0.075)	0.742 (0.075)
average of systems	0.414 (0.238)	0.599 (0.323)	0.600 (0.323)
worst system	0.249 (0.236)	0.376 (0.356)	0.377 (0.358)
best <u>baseline</u>	0.603 (0.603)	0.697 (0.697)	0.699 (0.699)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.871 (0.871)	0.925 (0.925)	0.926 (0.926)
average of systems	0.596 (0.494)	0.709 (0.573)	0.710 (0.574)
worst system	0.415 (0.246)	0.585 (0.582)	0.587 (0.584)
best <u>baseline</u>	0.535 (0.534)	0.591 (0.589)	0.592 (0.591)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.535 (0.176)	0.633 (0.208)	0.633 (0.208)
average of systems	0.535 (0.176)	0.633 (0.208)	0.633 (0.208)
worst system	0.535 (0.176)	0.633 (0.208)	0.633 (0.208)

Detailed results**trainable-adjectives**

The intersection of adjectives and trainable.

number of items in task:	1284
--------------------------	------

	all senses	main senses only
average polysemy:	6.927	4.536

	fine-grained	coarse-grained
average <u>entropy</u> :	1.700	1.237

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.964 (0.963)	0.971 (0.970)	0.972 (0.972)
best system	0.766 (0.738)	0.783 (0.754)	0.785 (0.756)
average of systems	0.603 (0.555)	0.641 (0.588)	0.657 (0.601)

worst system	0.181 (0.142)	0.312 (0.244)	0.323 (0.252)
best <u>baseline</u>	0.720 (0.719)	0.740 (0.739)	0.743 (0.743)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.549 (0.529)	0.569 (0.548)	0.612 (0.589)
average of systems	0.424 (0.354)	0.479 (0.398)	0.503 (0.418)
worst system	0.181 (0.142)	0.312 (0.244)	0.323 (0.252)
best <u>baseline</u>	0.660 (0.660)	0.677 (0.677)	0.700 (0.700)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.766 (0.738)	0.783 (0.754)	0.785 (0.756)
average of systems	0.692 (0.656)	0.722 (0.683)	0.734 (0.693)
worst system	0.589 (0.589)	0.631 (0.631)	0.634 (0.634)
best <u>baseline</u>	0.720 (0.719)	0.740 (0.739)	0.743 (0.743)

Detailed results**all-trainable-adjectives**

The intersection of all-adjectives and trainable.

number of items in task:	1628
--------------------------	------

	all senses	main senses only
average polysemy:	8.687	5.933

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.915	1.510

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.955 (0.954)	0.961 (0.960)	0.962 (0.961)
best system	0.740 (0.719)	0.753 (0.731)	0.755 (0.733)
average of systems	0.589 (0.542)	0.622 (0.570)	0.635 (0.581)
worst system	0.214 (0.176)	0.320 (0.263)	0.329 (0.270)
best <u>baseline</u>	0.686 (0.685)	0.703 (0.702)	0.706 (0.705)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.529 (0.500)	0.545 (0.515)	0.579 (0.547)
average of systems	0.418 (0.345)	0.466 (0.384)	0.485 (0.400)
worst system	0.214 (0.176)	0.320 (0.263)	0.329 (0.270)
best <u>baseline</u>	0.594 (0.594)	0.606 (0.606)	0.609 (0.609)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.740 (0.719)	0.753 (0.731)	0.755 (0.733)
average of systems	0.675 (0.640)	0.700 (0.663)	0.710 (0.672)
worst system	0.606 (0.606)	0.639 (0.639)	0.642 (0.642)
best <u>baseline</u>	0.686 (0.685)	0.703 (0.702)	0.706 (0.705)

Detailed results**all-untrainable-adjectives**

The intersection of all-adjectives and untrainable.

number of items in task:	122
--------------------------	-----

	all senses	main senses only
average polysemy:	5.000	5.000

	fine-grained	coarse-grained
average <u>entropy</u> :	1.224	1.224

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.985 (0.985)	0.985 (0.985)	0.985 (0.985)
best system	0.967 (0.967)	0.967 (0.967)	0.967 (0.967)
average of systems	0.808 (0.699)	0.838 (0.727)	0.838 (0.727)

worst system	0.617 (0.331)	0.617 (0.331)	0.617 (0.331)
best <u>baseline</u>	0.902 (0.902)	0.902 (0.902)	0.902 (0.902)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.892 (0.877)	0.892 (0.877)	0.892 (0.877)
average of systems	0.816 (0.646)	0.862 (0.687)	0.862 (0.687)
worst system	0.639 (0.566)	0.824 (0.730)	0.824 (0.730)
best <u>baseline</u>	0.902 (0.902)	0.902 (0.902)	0.902 (0.902)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.967 (0.967)	0.967 (0.967)	0.967 (0.967)
average of systems	0.801 (0.742)	0.819 (0.760)	0.819 (0.760)
worst system	0.617 (0.331)	0.617 (0.331)	0.617 (0.331)
best <u>baseline</u>	0.704 (0.693)	0.704 (0.693)	0.704 (0.693)

Detailed results**trainable-indeterminates**

The intersection of indeterminates and trainable.

number of items in task:	1462
--------------------------	------

	all senses	main senses only
average polysemy:	20.392	16.789

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.475	2.343

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.967 (0.965)	0.969 (0.967)	0.969 (0.967)
best system	0.776 (0.776)	0.782 (0.782)	0.784 (0.784)
average of systems	0.538 (0.475)	0.547 (0.482)	0.549 (0.484)
worst system	0.138 (0.009)	0.138 (0.009)	0.138 (0.009)
best <u>baseline</u>	0.656 (0.649)	0.658 (0.651)	0.661 (0.654)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.685 (0.611)	0.693 (0.619)	0.696 (0.621)
average of systems	0.335 (0.247)	0.346 (0.256)	0.348 (0.258)
worst system	0.138 (0.009)	0.138 (0.009)	0.138 (0.009)
best <u>baseline</u>	0.452 (0.451)	0.466 (0.465)	0.467 (0.466)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.776 (0.776)	0.782 (0.782)	0.784 (0.784)
average of systems	0.713 (0.678)	0.721 (0.686)	0.723 (0.688)
worst system	0.622 (0.622)	0.633 (0.633)	0.635 (0.635)
best <u>baseline</u>	0.656 (0.649)	0.658 (0.651)	0.661 (0.654)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.210 (0.035)	0.210 (0.035)	0.210 (0.035)
average of systems	0.210 (0.035)	0.210 (0.035)	0.210 (0.035)
worst system	0.210 (0.035)	0.210 (0.035)	0.210 (0.035)

Detailed results**trainable-determinates**

The intersection of determinates and trainable.

number of items in task:	5984
--------------------------	------

	all senses	main senses only
average polysemy:	8.442	5.146

	fine-grained	coarse-grained
average <u>entropy</u> :	1.837	1.358

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.962 (0.961)	0.967 (0.965)	0.969 (0.967)
best system	0.760 (0.751)	0.790 (0.780)	0.805 (0.795)
average of systems	0.555 (0.398)	0.644 (0.435)	0.681 (0.458)
worst system	0.180 (0.134)	0.315 (0.234)	0.349 (0.259)
best <u>baseline</u>	0.724 (0.723)	0.755 (0.754)	0.784 (0.783)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.549 (0.513)	0.753 (0.031)	0.800 (0.033)
average of systems	0.420 (0.225)	0.554 (0.270)	0.600 (0.293)
worst system	0.180 (0.134)	0.315 (0.234)	0.349 (0.259)
best <u>baseline</u>	0.573 (0.570)	0.610 (0.607)	0.631 (0.628)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.760 (0.751)	0.790 (0.780)	0.805 (0.795)
average of systems	0.661 (0.576)	0.705 (0.609)	0.737 (0.635)
worst system	0.438 (0.159)	0.564 (0.205)	0.612 (0.223)
best <u>baseline</u>	0.724 (0.723)	0.755 (0.754)	0.784 (0.783)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.649 (0.179)	0.733 (0.202)	0.767 (0.211)
average of systems	0.535 (0.112)	0.678 (0.134)	0.712 (0.140)
worst system	0.420 (0.045)	0.622 (0.066)	0.657 (0.070)

Detailed results**low-polysemy**

Items involving words whose polysemy is less than the median polysemy of 8.

number of items in task:	3771
--------------------------	------

	all senses	main senses only
average polysemy:	5.069	3.682

	fine-grained	coarse-grained
average <u>entropy</u> :	1.281	1.011

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.971 (0.970)	0.974 (0.973)	0.975 (0.973)
best system	0.839 (0.829)	0.857 (0.846)	0.865 (0.854)
average of systems	0.613 (0.407)	0.670 (0.431)	0.683 (0.439)
worst system	0.244 (0.178)	0.368 (0.268)	0.370 (0.270)

best <u>baseline</u>	0.780 (0.678)	0.798 (0.694)	0.810 (0.704)
----------------------	----------------------	----------------------	----------------------

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.614 (0.583)	0.661 (0.628)	0.683 (0.649)
average of systems	0.466 (0.248)	0.553 (0.281)	0.566 (0.289)
worst system	0.244 (0.178)	0.368 (0.268)	0.370 (0.270)
best <u>baseline</u>	0.675 (0.675)	0.706 (0.706)	0.731 (0.730)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.839 (0.829)	0.857 (0.846)	0.865 (0.854)
average of systems	0.725 (0.580)	0.756 (0.600)	0.768 (0.610)
worst system	0.522 (0.126)	0.584 (0.513)	0.590 (0.518)
best <u>baseline</u>	0.780 (0.678)	0.798 (0.694)	0.810 (0.704)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.724 (0.106)	0.782 (0.115)	0.806 (0.118)
average of systems	0.602 (0.080)	0.681 (0.090)	0.698 (0.093)
worst system	0.481 (0.054)	0.581 (0.066)	0.590 (0.067)

Detailed results

high-polysemy

Items involving words whose polysemy is equal to or greater than the median polysemy of 8.

number of items in task:	4677
--------------------------	------

	all senses	main senses only
average polysemy:	14.648	10.050

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	2.428	1.916

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.959 (0.957)	0.963 (0.962)	0.965 (0.964)
best system	0.725 (0.725)	0.760 (0.760)	0.783 (0.783)
average of systems	0.505 (0.352)	0.604 (0.393)	0.642 (0.416)
worst system	0.178 (0.149)	0.278 (0.233)	0.316 (0.265)
best <u>baseline</u>	0.624 (0.623)	0.658 (0.657)	0.689 (0.688)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.513 (0.467)	0.686 (0.038)	0.730 (0.040)
average of systems	0.363 (0.201)	0.512 (0.250)	0.558 (0.273)
worst system	0.178 (0.149)	0.278 (0.233)	0.316 (0.265)

best <u>baseline</u>	0.465 (0.462)	0.504 (0.501)	0.527 (0.524)
----------------------	----------------------	----------------------	----------------------

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.725 (0.725)	0.760 (0.760)	0.783 (0.783)
average of systems	0.615 (0.508)	0.671 (0.545)	0.705 (0.571)
worst system	0.388 (0.148)	0.572 (0.218)	0.622 (0.238)
best <u>baseline</u>	0.624 (0.623)	0.658 (0.657)	0.689 (0.688)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.611 (0.143)	0.709 (0.166)	0.748 (0.175)
average of systems	0.486 (0.100)	0.613 (0.124)	0.645 (0.131)
worst system	0.362 (0.057)	0.517 (0.082)	0.541 (0.086)

Detailed results**low-entropy**

Items involving words whose entropy is less than the median entropy of 1.85.

number of items in task:	3872
--------------------------	------

	all senses	main senses only
average polysemy:	7.309	5.472

	<u>fine-grained</u>	<u>coarse-grained</u>
average <u>entropy</u> :	1.116	0.846

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.979 (0.978)	0.981 (0.980)	0.981 (0.981)
best system	0.912 (0.912)	0.929 (0.918)	0.934 (0.923)
average of systems	0.644 (0.452)	0.727 (0.487)	0.742 (0.496)
worst system	0.267 (0.213)	0.410 (0.327)	0.418 (0.333)
best <u>baseline</u>	0.877 (0.713)	0.892 (0.725)	0.907 (0.738)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.728 (0.671)	0.765 (0.705)	0.787 (0.726)
average of systems	0.471 (0.282)	0.601 (0.333)	0.618 (0.342)
worst system	0.267 (0.213)	0.410 (0.327)	0.418 (0.333)
best <u>baseline</u>	0.696 (0.696)	0.728 (0.728)	0.749 (0.749)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.912 (0.912)	0.929 (0.918)	0.934 (0.923)
average of systems	0.784 (0.635)	0.826 (0.660)	0.840 (0.669)
worst system	0.483 (0.193)	0.664 (0.548)	0.669 (0.552)
best <u>baseline</u>	0.877 (0.713)	0.892 (0.725)	0.907 (0.738)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.728 (0.164)	0.803 (0.181)	0.822 (0.185)
average of systems	0.587 (0.120)	0.696 (0.141)	0.707 (0.144)
worst system	0.446 (0.077)	0.589 (0.102)	0.592 (0.103)

Detailed results**high-entropy**

Items involving words whose entropy is equal to or greater than the median entropy of 1.85.

number of items in task:	4576
--------------------------	------

	all senses	main senses only
average polysemy:	12.963	8.675

	fine-grained	coarse-grained
average entropy:	2.593	2.075

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.953 (0.950)	0.957 (0.955)	0.960 (0.957)
best system	0.653 (0.607)	0.702 (0.020)	0.752 (0.021)
average of systems	0.471 (0.312)	0.552 (0.344)	0.596 (0.368)

worst system	0.152 (0.119)	0.233 (0.182)	0.270 (0.211)
best <u>baseline</u>	0.579 (0.578)	0.615 (0.614)	0.643 (0.642)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.450 (0.013)	0.702 (0.020)	0.752 (0.021)
average of systems	0.358 (0.170)	0.478 (0.205)	0.532 (0.227)
worst system	0.152 (0.119)	0.233 (0.182)	0.270 (0.211)
best <u>baseline</u>	0.454 (0.452)	0.495 (0.492)	0.519 (0.516)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.653 (0.607)	0.690 (0.690)	0.718 (0.717)
average of systems	0.559 (0.460)	0.606 (0.493)	0.641 (0.519)
worst system	0.367 (0.093)	0.473 (0.119)	0.534 (0.135)
best <u>baseline</u>	0.579 (0.578)	0.615 (0.614)	0.643 (0.642)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.561 (0.095)	0.654 (0.111)	0.706 (0.120)
average of systems	0.456 (0.067)	0.564 (0.082)	0.610 (0.088)
worst system	0.350 (0.038)	0.474 (0.052)	0.514 (0.056)

Detailed results

accident-n

number of items in task:	267
--------------------------	-----

	all senses	main senses only
polysemy:	8	2

	fine-grained	coarse-grained
entropy:	1.430	0.571

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.987 (0.987)	0.988 (0.988)	0.991 (0.991)
best system	0.944 (0.944)	0.954 (0.954)	0.987 (0.734)
average of systems	0.658 (0.594)	0.803 (0.689)	0.878 (0.752)
worst system	0.229 (0.064)	0.489 (0.488)	0.556 (0.554)
best <u>baseline</u>	0.789 (0.783)	0.828 (0.828)	0.963 (0.963)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.856 (0.801)	0.884 (0.828)	0.972 (0.525)
average of systems	0.494 (0.389)	0.725 (0.533)	0.809 (0.594)
worst system	0.247 (0.029)	0.489 (0.488)	0.556 (0.554)
best <u>baseline</u>	0.753 (0.753)	0.789 (0.789)	0.933 (0.933)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.944 (0.944)	0.954 (0.954)	0.987 (0.734)
average of systems	0.819 (0.802)	0.876 (0.857)	0.950 (0.929)
worst system	0.423 (0.423)	0.720 (0.720)	0.808 (0.808)
best <u>baseline</u>	0.789 (0.783)	0.828 (0.828)	0.963 (0.963)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.644 (0.484)	0.767 (0.576)	0.823 (0.618)
average of systems	0.437 (0.274)	0.715 (0.381)	0.752 (0.404)
worst system	0.229 (0.064)	0.662 (0.186)	0.680 (0.191)

Detailed results**behaviour-n**

number of items in task:	279
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
entropy:	0.390	0.295

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.973 (0.973)	0.973 (0.973)	0.973 (0.973)
best system	0.964 (0.964)	0.964 (0.964)	0.964 (0.964)
average of systems	0.771 (0.676)	0.909 (0.759)	0.909 (0.759)
worst system	0.263 (0.251)	0.673 (0.642)	0.673 (0.642)
best <u>baseline</u>	0.946 (0.946)	0.961 (0.961)	0.961 (0.961)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.924 (0.910)	0.959 (0.928)	0.959 (0.928)
average of systems	0.585 (0.414)	0.853 (0.576)	0.853 (0.576)
worst system	0.263 (0.251)	0.673 (0.642)	0.673 (0.642)
best <u>baseline</u>	0.946 (0.946)	0.961 (0.961)	0.961 (0.961)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.964 (0.964)	0.964 (0.964)	0.964 (0.964)
average of systems	0.918 (0.889)	0.945 (0.915)	0.945 (0.915)
worst system	0.575 (0.575)	0.796 (0.796)	0.796 (0.796)
best <u>baseline</u>	0.946 (0.946)	0.961 (0.961)	0.961 (0.961)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.777 (0.747)	0.951 (0.276)	0.951 (0.276)
average of systems	0.626 (0.442)	0.914 (0.559)	0.914 (0.559)
worst system	0.475 (0.138)	0.877 (0.843)	0.877 (0.843)

Detailed results**bet-n**

number of items in task:	274
--------------------------	-----

	all senses	main senses only
polysemy:	15	9

	fine-grained	coarse-grained
<u>entropy</u> :	3.200	2.563

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.970 (0.970)	0.980 (0.980)	0.980 (0.980)
best system	0.661 (0.661)	0.807 (0.807)	0.869 (0.869)
average of systems	0.471 (0.376)	0.548 (0.434)	0.593 (0.474)
worst system	0.132 (0.131)	0.194 (0.135)	0.194 (0.135)
best <u>baseline</u>	0.547 (0.547)	0.652 (0.621)	0.782 (0.745)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.629 (0.080)	0.800 (0.102)	0.800 (0.102)
average of systems	0.394 (0.234)	0.488 (0.288)	0.533 (0.326)
worst system	0.132 (0.131)	0.194 (0.135)	0.194 (0.135)
best <u>baseline</u>	0.406 (0.400)	0.426 (0.404)	0.452 (0.429)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.661 (0.661)	0.807 (0.807)	0.869 (0.869)
average of systems	0.515 (0.500)	0.583 (0.565)	0.635 (0.613)
worst system	0.354 (0.354)	0.413 (0.411)	0.419 (0.418)
best <u>baseline</u>	0.547 (0.547)	0.652 (0.621)	0.782 (0.745)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.615 (0.263)	0.675 (0.288)	0.675 (0.288)
average of systems	0.550 (0.274)	0.603 (0.299)	0.616 (0.308)
worst system	0.485 (0.285)	0.531 (0.311)	0.557 (0.327)

Detailed results**disability-n**

number of items in task:	160
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
entropy:	1.053	0.457

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.959 (0.959)	0.959 (0.959)	0.959 (0.959)
best system	0.900 (0.900)	0.942 (0.906)	0.942 (0.906)
average of systems	0.735 (0.555)	0.821 (0.625)	0.821 (0.625)
worst system	0.408 (0.400)	0.484 (0.235)	0.484 (0.235)
best <u>baseline</u>	0.800 (0.800)	0.938 (0.938)	0.938 (0.938)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.870 (0.837)	0.942 (0.906)	0.942 (0.906)
average of systems	0.718 (0.487)	0.806 (0.556)	0.806 (0.556)
worst system	0.408 (0.400)	0.484 (0.235)	0.484 (0.235)
best <u>baseline</u>	0.800 (0.800)	0.938 (0.938)	0.938 (0.938)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.900 (0.900)	0.938 (0.938)	0.938 (0.938)
average of systems	0.745 (0.701)	0.829 (0.782)	0.829 (0.782)
worst system	0.459 (0.450)	0.567 (0.556)	0.567 (0.556)
best <u>baseline</u>	0.755 (0.750)	0.906 (0.900)	0.906 (0.900)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.815 (0.138)	0.889 (0.150)	0.889 (0.150)
average of systems	0.815 (0.138)	0.889 (0.150)	0.889 (0.150)
worst system	0.815 (0.138)	0.889 (0.150)	0.889 (0.150)

Detailed results**excess-n**

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	8	3

	fine-grained	coarse-grained
entropy:	2.387	1.199

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.973 (0.968)	0.973 (0.968)	0.973 (0.968)
best system	0.882 (0.882)	0.899 (0.899)	0.919 (0.919)
average of systems	0.495 (0.478)	0.587 (0.551)	0.692 (0.641)
worst system	0.057 (0.013)	0.236 (0.129)	0.397 (0.216)
best <u>baseline</u>	0.800 (0.796)	0.820 (0.816)	0.881 (0.876)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.456 (0.414)	0.547 (0.497)	0.822 (0.747)
average of systems	0.267 (0.248)	0.361 (0.337)	0.530 (0.489)
worst system	0.168 (0.167)	0.236 (0.129)	0.397 (0.216)
best <u>baseline</u>	0.661 (0.661)	0.747 (0.747)	0.747 (0.747)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.882 (0.882)	0.899 (0.899)	0.919 (0.919)
average of systems	0.662 (0.652)	0.726 (0.712)	0.801 (0.775)
worst system	0.263 (0.161)	0.396 (0.243)	0.545 (0.545)
best <u>baseline</u>	0.800 (0.796)	0.820 (0.816)	0.881 (0.876)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.532 (0.466)	0.663 (0.580)	0.723 (0.633)
average of systems	0.294 (0.240)	0.546 (0.337)	0.609 (0.371)
worst system	0.057 (0.013)	0.429 (0.095)	0.496 (0.109)

Detailed results**float-n**

number of items in task:	75
--------------------------	----

	all senses	main senses only
polysemy:	12	8

	fine-grained	coarse-grained
entropy:	2.340	2.042

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.980 (0.980)	0.980 (0.980)	0.993 (0.993)
best system	0.813 (0.813)	0.813 (0.813)	0.840 (0.840)
average of systems	0.367 (0.352)	0.402 (0.379)	0.493 (0.453)
worst system	0.076 (0.067)	0.076 (0.067)	0.121 (0.107)
best <u>baseline</u>	0.720 (0.720)	0.720 (0.720)	0.733 (0.733)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.347 (0.347)	0.347 (0.347)	0.411 (0.189)
average of systems	0.165 (0.145)	0.200 (0.175)	0.293 (0.254)
worst system	0.076 (0.067)	0.076 (0.067)	0.121 (0.107)
best <u>baseline</u>	0.267 (0.260)	0.281 (0.273)	0.520 (0.520)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.813 (0.813)	0.813 (0.813)	0.840 (0.840)
average of systems	0.502 (0.498)	0.534 (0.523)	0.619 (0.597)
worst system	0.107 (0.107)	0.271 (0.271)	0.357 (0.357)
best <u>baseline</u>	0.720 (0.720)	0.720 (0.720)	0.733 (0.733)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.167 (0.053)	0.236 (0.076)	0.375 (0.120)
average of systems	0.167 (0.053)	0.236 (0.076)	0.375 (0.120)
worst system	0.167 (0.053)	0.236 (0.076)	0.375 (0.120)

Detailed results

giant-n

number of items in task:	118
--------------------------	-----

	all senses	main senses only
polysemy:	7	3

	fine-grained	coarse-grained
entropy:	2.054	1.239

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.975 (0.975)	0.975 (0.975)	0.992 (0.992)
best system	0.856 (0.856)	0.890 (0.890)	0.983 (0.983)
average of systems	0.487 (0.447)	0.567 (0.512)	0.710 (0.630)
worst system	0.085 (0.085)	0.085 (0.085)	0.085 (0.085)
best <u>baseline</u>	0.763 (0.763)	0.839 (0.839)	0.983 (0.983)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.492 (0.492)	0.568 (0.568)	0.720 (0.720)
average of systems	0.271 (0.235)	0.365 (0.307)	0.525 (0.435)
worst system	0.085 (0.085)	0.085 (0.085)	0.085 (0.085)
best <u>baseline</u>	0.534 (0.534)	0.571 (0.571)	0.720 (0.720)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.856 (0.856)	0.890 (0.890)	0.983 (0.983)
average of systems	0.626 (0.605)	0.701 (0.671)	0.835 (0.790)
worst system	0.274 (0.274)	0.330 (0.330)	0.460 (0.460)
best <u>baseline</u>	0.763 (0.763)	0.839 (0.839)	0.983 (0.983)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.503 (0.414)	0.557 (0.459)	0.695 (0.572)
average of systems	0.407 (0.245)	0.465 (0.275)	0.606 (0.350)
worst system	0.310 (0.076)	0.374 (0.092)	0.517 (0.127)

Detailed results**knee-n**

number of items in task:	251
--------------------------	-----

	all senses	main senses only
polysemy:	22	12

	fine-grained	coarse-grained
entropy:	2.484	1.463

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.988 (0.988)	0.988 (0.988)	0.988 (0.988)
best system	0.824 (0.821)	0.958 (0.092)	1.000 (0.096)
average of systems	0.519 (0.466)	0.708 (0.557)	0.793 (0.626)
worst system	0.000 (0.000)	0.363 (0.131)	0.462 (0.167)
best <u>baseline</u>	0.665 (0.665)	0.819 (0.803)	0.861 (0.861)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.661 (0.582)	0.958 (0.092)	1.000 (0.096)
average of systems	0.292 (0.243)	0.623 (0.363)	0.736 (0.447)
worst system	0.000 (0.000)	0.363 (0.131)	0.462 (0.167)
best <u>baseline</u>	0.578 (0.578)	0.656 (0.656)	0.833 (0.833)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.824 (0.821)	0.877 (0.877)	0.916 (0.916)
average of systems	0.689 (0.667)	0.765 (0.740)	0.837 (0.806)
worst system	0.394 (0.391)	0.653 (0.648)	0.756 (0.750)
best <u>baseline</u>	0.665 (0.665)	0.819 (0.803)	0.861 (0.861)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.720 (0.447)	0.768 (0.477)	0.800 (0.497)
average of systems	0.518 (0.266)	0.747 (0.335)	0.780 (0.350)
worst system	0.316 (0.084)	0.725 (0.193)	0.761 (0.203)

Detailed results**onion-n**

number of items in task:	214
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	fine-grained	coarse-grained
entropy:	0.862	0.862

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.951 (0.951)	0.951 (0.951)	0.951 (0.951)
best system	1.000 (0.019)	1.000 (0.019)	1.000 (0.019)
average of systems	0.795 (0.639)	0.795 (0.639)	0.795 (0.639)
worst system	0.434 (0.355)	0.434 (0.355)	0.434 (0.355)
best <u>baseline</u>	0.911 (0.911)	0.911 (0.911)	0.911 (0.911)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	1.000 (0.019)	1.000 (0.019)	1.000 (0.019)
average of systems	0.710 (0.439)	0.710 (0.439)	0.710 (0.439)
worst system	0.434 (0.355)	0.434 (0.355)	0.434 (0.355)
best <u>baseline</u>	0.911 (0.911)	0.911 (0.911)	0.911 (0.911)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.925 (0.925)	0.925 (0.925)	0.925 (0.925)
average of systems	0.857 (0.817)	0.857 (0.817)	0.857 (0.817)
worst system	0.696 (0.696)	0.696 (0.696)	0.696 (0.696)
best <u>baseline</u>	0.909 (0.841)	0.909 (0.841)	0.909 (0.841)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.858 (0.269)	0.858 (0.269)	0.858 (0.269)
average of systems	0.810 (0.468)	0.810 (0.468)	0.810 (0.468)
worst system	0.762 (0.668)	0.762 (0.668)	0.762 (0.668)

Detailed results**promise-n**

number of items in task:	113
--------------------------	-----

	all senses	main senses only
polysemy:	8	4

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	1.851	0.963

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.965 (0.965)	0.965 (0.965)	0.965 (0.965)
best system	0.867 (0.867)	0.889 (0.889)	0.920 (0.920)
average of systems	0.566 (0.497)	0.620 (0.541)	0.705 (0.598)
worst system	0.000 (0.000)	0.124 (0.124)	0.133 (0.133)
best <u>baseline</u>	0.717 (0.717)	0.746 (0.746)	0.832 (0.832)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.630 (0.602)	0.662 (0.633)	0.759 (0.726)
average of systems	0.413 (0.294)	0.479 (0.338)	0.588 (0.390)
worst system	0.000 (0.000)	0.124 (0.124)	0.133 (0.133)
best <u>baseline</u>	0.708 (0.708)	0.737 (0.737)	0.823 (0.823)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.867 (0.867)	0.889 (0.889)	0.920 (0.920)
average of systems	0.698 (0.677)	0.737 (0.716)	0.804 (0.779)
worst system	0.362 (0.362)	0.378 (0.378)	0.419 (0.419)
best <u>baseline</u>	0.717 (0.717)	0.746 (0.746)	0.832 (0.832)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.606 (0.538)	0.748 (0.664)	0.786 (0.698)
average of systems	0.463 (0.336)	0.555 (0.407)	0.637 (0.451)
worst system	0.319 (0.133)	0.362 (0.150)	0.489 (0.204)

Detailed results**rabbit-n**

number of items in task:	221
--------------------------	-----

	all senses	main senses only
polysemy:	8	6

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy</u> :	0.748	0.522

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.948 (0.948)	0.950 (0.950)	0.952 (0.952)
best system	0.946 (0.946)	0.955 (0.955)	0.964 (0.964)
average of systems	0.499 (0.417)	0.809 (0.618)	0.813 (0.621)
worst system	0.000 (0.000)	0.427 (0.425)	0.427 (0.425)
best <u>baseline</u>	0.919 (0.919)	0.928 (0.928)	0.937 (0.937)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.921 (0.787)	0.938 (0.722)	0.941 (0.724)
average of systems	0.385 (0.316)	0.770 (0.531)	0.773 (0.533)
worst system	0.000 (0.000)	0.427 (0.425)	0.427 (0.425)
best <u>baseline</u>	0.919 (0.919)	0.928 (0.928)	0.937 (0.937)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.946 (0.946)	0.955 (0.955)	0.964 (0.964)
average of systems	0.647 (0.588)	0.846 (0.787)	0.852 (0.792)
worst system	0.372 (0.371)	0.616 (0.209)	0.619 (0.210)
best <u>baseline</u>	0.600 (0.600)	0.624 (0.624)	0.631 (0.631)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.500 (0.122)	0.907 (0.222)	0.907 (0.222)
average of systems	0.500 (0.122)	0.907 (0.222)	0.907 (0.222)
worst system	0.500 (0.122)	0.907 (0.222)	0.907 (0.222)

Detailed results**sack-n**

number of items in task:	82
--------------------------	----

	all senses	main senses only
polysemy:	11	9

	fine-grained	coarse-grained
entropy:	1.772	1.667

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
best system	0.878 (0.878)	0.878 (0.878)	0.878 (0.878)
average of systems	0.594 (0.491)	0.609 (0.505)	0.609 (0.505)
worst system	0.204 (0.067)	0.204 (0.067)	0.204 (0.067)
best <u>baseline</u>	0.889 (0.878)	0.889 (0.878)	0.889 (0.878)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.812 (0.476)	0.812 (0.476)	0.812 (0.476)
average of systems	0.522 (0.316)	0.543 (0.336)	0.543 (0.336)
worst system	0.218 (0.207)	0.328 (0.328)	0.328 (0.328)
best <u>baseline</u>	0.524 (0.524)	0.524 (0.524)	0.524 (0.524)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.878 (0.878)	0.878 (0.878)	0.878 (0.878)
average of systems	0.672 (0.650)	0.682 (0.659)	0.682 (0.659)
worst system	0.328 (0.328)	0.339 (0.339)	0.339 (0.339)
best <u>baseline</u>	0.889 (0.878)	0.889 (0.878)	0.889 (0.878)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.693 (0.581)	0.732 (0.614)	0.732 (0.614)
average of systems	0.449 (0.324)	0.468 (0.340)	0.468 (0.340)
worst system	0.204 (0.067)	0.204 (0.067)	0.204 (0.067)

Detailed results

scrap-n

number of items in task:	156
--------------------------	-----

	all senses	main senses only
polysemy:	14	8

	fine-grained	coarse-grained
entropy:	2.839	1.999

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.965 (0.965)	0.974 (0.974)	0.978 (0.978)
best system	0.686 (0.686)	0.966 (0.179)	0.966 (0.179)
average of systems	0.490 (0.357)	0.641 (0.456)	0.696 (0.496)
worst system	0.097 (0.096)	0.161 (0.160)	0.161 (0.160)
best <u>baseline</u>	0.622 (0.622)	0.760 (0.760)	0.795 (0.795)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.586 (0.109)	0.966 (0.179)	0.966 (0.179)
average of systems	0.474 (0.223)	0.657 (0.297)	0.718 (0.327)
worst system	0.097 (0.096)	0.161 (0.160)	0.161 (0.160)
best <u>baseline</u>	0.622 (0.622)	0.760 (0.760)	0.795 (0.795)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.686 (0.686)	0.833 (0.833)	0.865 (0.865)
average of systems	0.525 (0.488)	0.659 (0.611)	0.713 (0.662)
worst system	0.250 (0.250)	0.256 (0.256)	0.263 (0.263)
best <u>baseline</u>	0.583 (0.583)	0.708 (0.708)	0.782 (0.782)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.489 (0.314)	0.661 (0.423)	0.698 (0.448)
average of systems	0.352 (0.181)	0.466 (0.242)	0.499 (0.257)
worst system	0.214 (0.048)	0.271 (0.061)	0.300 (0.067)

Detailed results**shirt-n**

number of items in task:	184
--------------------------	-----

	<u>all senses</u>	<u>main senses only</u>
<u>polysemy</u> :	8	6

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy</u> :	1.778	1.235

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.992 (0.992)	0.995 (0.995)	0.997 (0.997)
best system	0.967 (0.967)	0.978 (0.978)	0.991 (0.592)
average of systems	0.701 (0.554)	0.782 (0.613)	0.834 (0.653)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.858 (0.821)	0.920 (0.880)	0.983 (0.940)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.832 (0.337)	0.940 (0.557)	0.991 (0.592)
average of systems	0.601 (0.376)	0.712 (0.444)	0.771 (0.481)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.821 (0.821)	0.899 (0.899)	0.940 (0.940)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.967 (0.967)	0.978 (0.978)	0.989 (0.989)
average of systems	0.777 (0.723)	0.840 (0.780)	0.890 (0.826)
worst system	0.441 (0.441)	0.504 (0.504)	0.555 (0.555)
best <u>baseline</u>	0.858 (0.821)	0.920 (0.880)	0.983 (0.940)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.727 (0.574)	0.806 (0.636)	0.841 (0.664)
average of systems	0.691 (0.338)	0.748 (0.373)	0.783 (0.389)
worst system	0.655 (0.103)	0.690 (0.109)	0.724 (0.114)

Detailed results**steering-n**

number of items in task:	176
--------------------------	-----

	all senses	main senses only
polysemy:	5	4

	fine-grained	coarse-grained
entropy:	1.712	1.319

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.989 (0.989)	0.989 (0.989)	0.989 (0.989)
best system	0.949 (0.949)	0.972 (0.972)	0.972 (0.972)
average of systems	0.405 (0.318)	0.428 (0.335)	0.428 (0.335)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.716 (0.716)	0.775 (0.775)	0.775 (0.775)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.685 (0.293)	0.746 (0.319)	0.746 (0.319)
average of systems	0.184 (0.129)	0.203 (0.141)	0.203 (0.141)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.716 (0.716)	0.775 (0.775)	0.775 (0.775)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.949 (0.949)	0.972 (0.972)	0.972 (0.972)
average of systems	0.604 (0.527)	0.635 (0.552)	0.635 (0.552)
worst system	0.043 (0.034)	0.043 (0.034)	0.043 (0.034)
best <u>baseline</u>	0.716 (0.716)	0.744 (0.744)	0.744 (0.744)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.774 (0.369)	0.774 (0.369)	0.774 (0.369)
average of systems	0.774 (0.369)	0.774 (0.369)	0.774 (0.369)
worst system	0.774 (0.369)	0.774 (0.369)	0.774 (0.369)

Detailed results

amaze-v

number of items in task:	70
--------------------------	----

	all senses	main senses only
polysemy:	1	1

	fine-grained	coarse-grained
entropy:	0.000	0.000

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
best system	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
average of systems	0.945 (0.874)	0.945 (0.874)	0.945 (0.874)
worst system	0.765 (0.743)	0.765 (0.743)	0.765 (0.743)
best <u>baseline</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	1.000 (0.971)	1.000 (0.971)	1.000 (0.971)
average of systems	0.923 (0.771)	0.923 (0.771)	0.923 (0.771)
worst system	0.807 (0.357)	0.807 (0.357)	0.807 (0.357)
best <u>baseline</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
average of systems	0.956 (0.925)	0.956 (0.925)	0.956 (0.925)
worst system	0.765 (0.743)	0.765 (0.743)	0.765 (0.743)
best <u>baseline</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)

Detailed results**bet-v**

number of items in task:	117
--------------------------	-----

	all senses	main senses only
polysemy:	9	4

	fine-grained	coarse-grained
entropy:	2.349	1.581

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.924 (0.916)	0.932 (0.925)	0.932 (0.925)
best system	0.769 (0.769)	0.778 (0.778)	0.812 (0.812)
average of systems	0.345 (0.336)	0.400 (0.385)	0.508 (0.481)
worst system	0.009 (0.009)	0.009 (0.009)	0.009 (0.009)

best <u>baseline</u>	0.714 (0.714)	0.726 (0.726)	0.803 (0.803)
----------------------	----------------------	----------------------	----------------------

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.509 (0.479)	0.518 (0.487)	0.545 (0.513)
average of systems	0.207 (0.190)	0.261 (0.238)	0.350 (0.317)
worst system	0.034 (0.034)	0.078 (0.077)	0.129 (0.128)
best <u>baseline</u>	0.714 (0.714)	0.726 (0.726)	0.803 (0.803)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.769 (0.769)	0.778 (0.778)	0.812 (0.812)
average of systems	0.414 (0.409)	0.470 (0.458)	0.587 (0.563)
worst system	0.009 (0.009)	0.009 (0.009)	0.009 (0.009)
best <u>baseline</u>	0.692 (0.692)	0.701 (0.701)	0.795 (0.795)

Detailed results**bother-v**

number of items in task:	209
--------------------------	-----

	all senses	main senses only
polysemy:	8	6

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	2.168	1.837

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.976 (0.976)	0.976 (0.976)	0.976 (0.976)
best system	0.852 (0.852)	0.871 (0.871)	0.871 (0.871)
average of systems	0.595 (0.567)	0.623 (0.592)	0.623 (0.592)
worst system	0.142 (0.134)	0.168 (0.158)	0.168 (0.158)
best <u>baseline</u>	0.632 (0.617)	0.637 (0.622)	0.637 (0.622)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.491 (0.388)	0.545 (0.431)	0.545 (0.431)
average of systems	0.379 (0.322)	0.424 (0.361)	0.424 (0.361)
worst system	0.142 (0.134)	0.168 (0.158)	0.168 (0.158)
best <u>baseline</u>	0.415 (0.405)	0.467 (0.467)	0.467 (0.467)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.852 (0.852)	0.871 (0.871)	0.871 (0.871)
average of systems	0.703 (0.689)	0.722 (0.708)	0.722 (0.708)
worst system	0.512 (0.512)	0.541 (0.541)	0.541 (0.541)
best <u>baseline</u>	0.632 (0.617)	0.637 (0.622)	0.637 (0.622)

Detailed results**bury-v**

number of items in task:	201
--------------------------	-----

	all senses	main senses only
polysemy:	14	6

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	2.759	2.401

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.928 (0.923)	0.930 (0.925)	0.933 (0.928)
best system	0.562 (0.562)	0.570 (0.570)	0.587 (0.587)
average of systems	0.396 (0.366)	0.421 (0.388)	0.436 (0.402)
worst system	0.241 (0.224)	0.270 (0.270)	0.289 (0.289)
best <u>baseline</u>	0.552 (0.552)	0.557 (0.557)	0.567 (0.567)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.300 (0.124)	0.318 (0.296)	0.331 (0.136)
average of systems	0.265 (0.205)	0.298 (0.234)	0.312 (0.244)

worst system	0.241 (0.224)	0.270 (0.270)	0.289 (0.289)
best <u>baseline</u>	0.365 (0.365)	0.383 (0.383)	0.384 (0.384)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.562 (0.562)	0.570 (0.570)	0.587 (0.587)
average of systems	0.449 (0.430)	0.469 (0.449)	0.486 (0.465)
worst system	0.398 (0.398)	0.417 (0.417)	0.423 (0.423)
best <u>baseline</u>	0.552 (0.552)	0.557 (0.557)	0.567 (0.567)

Detailed results**calculate-v**

number of items in task:	218
--------------------------	-----

	all senses	main senses only
polysemy:	5	3

	fine-grained	coarse-grained
entropy:	0.982	0.864

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.954 (0.950)	0.959 (0.954)	0.959 (0.954)
best system	0.922 (0.922)	0.922 (0.922)	0.922 (0.922)
average of systems	0.678 (0.635)	0.694 (0.649)	0.694 (0.649)
worst system	0.272 (0.271)	0.272 (0.271)	0.272 (0.271)
best <u>baseline</u>	0.904 (0.904)	0.904 (0.904)	0.904 (0.904)

[A systems](#)

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.512 (0.509)	0.614 (0.518)	0.614 (0.518)
average of systems	0.459 (0.382)	0.503 (0.421)	0.503 (0.421)
worst system	0.333 (0.257)	0.333 (0.257)	0.333 (0.257)
best <u>baseline</u>	0.493 (0.493)	0.493 (0.493)	0.493 (0.493)

[S systems](#)

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.922 (0.922)	0.922 (0.922)	0.922 (0.922)
average of systems	0.788 (0.761)	0.790 (0.763)	0.790 (0.763)
worst system	0.272 (0.271)	0.272 (0.271)	0.272 (0.271)
best <u>baseline</u>	0.904 (0.904)	0.904 (0.904)	0.904 (0.904)

[Detailed results](#)

consume-v

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	6	4

	fine-grained	coarse-grained
entropy:	2.218	1.677

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.944 (0.939)	0.955 (0.950)	0.958 (0.953)
best system	0.535 (0.532)	0.589 (0.586)	0.632 (0.629)
average of systems	0.412 (0.378)	0.489 (0.448)	0.532 (0.486)
worst system	0.189 (0.188)	0.400 (0.398)	0.416 (0.414)
best <u>baseline</u>	0.546 (0.543)	0.608 (0.605)	0.654 (0.651)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.397 (0.392)	0.454 (0.215)	0.513 (0.243)
average of systems	0.332 (0.274)	0.437 (0.366)	0.474 (0.396)
worst system	0.189 (0.188)	0.400 (0.398)	0.416 (0.414)
best <u>baseline</u>	0.416 (0.414)	0.500 (0.497)	0.535 (0.532)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.535 (0.532)	0.589 (0.586)	0.632 (0.629)
average of systems	0.452 (0.430)	0.515 (0.488)	0.562 (0.531)
worst system	0.373 (0.371)	0.438 (0.435)	0.489 (0.484)
best <u>baseline</u>	0.546 (0.543)	0.608 (0.605)	0.654 (0.651)

Detailed results**derive-v**

number of items in task:	217
--------------------------	-----

	all senses	main senses only
polysemy:	6	4

	fine-grained	coarse-grained
entropy:	1.955	1.731

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.965 (0.961)	0.965 (0.961)	0.965 (0.961)
best system	0.671 (0.668)	0.671 (0.668)	0.673 (0.673)
average of systems	0.504 (0.467)	0.523 (0.484)	0.530 (0.490)

worst system	0.161 (0.161)	0.249 (0.249)	0.258 (0.258)
best <u>baseline</u>	0.588 (0.585)	0.588 (0.585)	0.588 (0.585)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.526 (0.475)	0.526 (0.475)	0.526 (0.475)
average of systems	0.387 (0.320)	0.423 (0.353)	0.433 (0.361)
worst system	0.161 (0.161)	0.249 (0.249)	0.258 (0.258)
best <u>baseline</u>	0.530 (0.530)	0.530 (0.530)	0.530 (0.530)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.671 (0.668)	0.671 (0.668)	0.673 (0.673)
average of systems	0.563 (0.541)	0.573 (0.550)	0.578 (0.555)
worst system	0.486 (0.484)	0.498 (0.495)	0.509 (0.507)
best <u>baseline</u>	0.588 (0.585)	0.588 (0.585)	0.588 (0.585)

Detailed results**float-v**

number of items in task:	229
--------------------------	-----

	all senses	main senses only
polysemy:	16	11

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	3.333	2.632

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.927 (0.923)	0.938 (0.934)	0.943 (0.939)
best system	0.555 (0.555)	0.633 (0.633)	0.655 (0.655)
average of systems	0.369 (0.339)	0.419 (0.383)	0.442 (0.404)
worst system	0.021 (0.017)	0.037 (0.031)	0.042 (0.035)
best <u>baseline</u>	0.524 (0.524)	0.579 (0.579)	0.616 (0.616)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.394 (0.204)	0.471 (0.244)	0.507 (0.263)
average of systems	0.276 (0.225)	0.327 (0.267)	0.356 (0.291)
worst system	0.021 (0.017)	0.037 (0.031)	0.042 (0.035)
best <u>baseline</u>	0.403 (0.400)	0.467 (0.463)	0.502 (0.498)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.555 (0.555)	0.633 (0.633)	0.655 (0.655)
average of systems	0.422 (0.402)	0.471 (0.448)	0.489 (0.466)
worst system	0.279 (0.279)	0.285 (0.285)	0.288 (0.288)

best baseline	0.524 (0.524)	0.579 (0.579)	0.616 (0.616)
---------------	----------------------	----------------------	----------------------

Detailed results**invade-v**

number of items in task:	207
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	fine-grained	coarse-grained
entropy:	2.195	1.518

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.921 (0.912)	0.922 (0.913)	0.924 (0.915)
best system	0.598 (0.219)	0.645 (0.645)	0.676 (0.676)
average of systems	0.477 (0.422)	0.559 (0.489)	0.590 (0.518)
worst system	0.119 (0.034)	0.364 (0.104)	0.373 (0.106)
best baseline	0.570 (0.570)	0.643 (0.643)	0.686 (0.686)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.598 (0.219)	0.622 (0.227)	0.645 (0.236)
average of systems	0.411 (0.303)	0.528 (0.378)	0.550 (0.395)
worst system	0.119 (0.034)	0.364 (0.104)	0.373 (0.106)
best <u>baseline</u>	0.420 (0.418)	0.495 (0.493)	0.517 (0.514)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.580 (0.580)	0.645 (0.645)	0.676 (0.676)
average of systems	0.511 (0.482)	0.575 (0.545)	0.611 (0.579)
worst system	0.454 (0.454)	0.514 (0.514)	0.553 (0.551)
best <u>baseline</u>	0.570 (0.570)	0.643 (0.643)	0.686 (0.686)

Detailed results**promise-v**

number of items in task:	224
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	0.982	0.812

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.953 (0.953)	0.962 (0.962)	0.962 (0.962)
best system	0.906 (0.906)	0.913 (0.913)	0.920 (0.920)
average of systems	0.699 (0.655)	0.745 (0.700)	0.753 (0.708)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.862 (0.862)	0.873 (0.873)	0.884 (0.884)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.731 (0.728)	0.772 (0.672)	0.785 (0.683)
average of systems	0.563 (0.483)	0.681 (0.598)	0.691 (0.607)
worst system	0.058 (0.058)	0.473 (0.473)	0.478 (0.478)
best <u>baseline</u>	0.862 (0.862)	0.873 (0.873)	0.884 (0.884)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.906 (0.906)	0.913 (0.913)	0.920 (0.920)
average of systems	0.766 (0.741)	0.777 (0.751)	0.785 (0.758)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.857 (0.857)	0.868 (0.868)	0.879 (0.879)

[Detailed results](#)

sack-v.

number of items in task:	178
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	0.132	0.132

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.994 (0.994)	0.994 (0.994)	0.994 (0.994)
best system	0.982 (0.646)	0.982 (0.646)	0.982 (0.646)
average of systems	0.679 (0.646)	0.679 (0.646)	0.679 (0.646)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.980 (0.980)	0.980 (0.980)	0.980 (0.980)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.907 (0.826)	0.907 (0.826)	0.907 (0.826)
average of systems	0.301 (0.272)	0.301 (0.272)	0.301 (0.272)
worst system	0.006 (0.006)	0.006 (0.006)	0.006 (0.006)
best <u>baseline</u>	0.978 (0.978)	0.978 (0.978)	0.978 (0.978)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.982 (0.646)	0.982 (0.646)	0.982 (0.646)
average of systems	0.868 (0.833)	0.868 (0.834)	0.868 (0.834)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.980 (0.980)	0.980 (0.980)	0.980 (0.980)

Detailed results**scrap-v**

number of items in task:	186
--------------------------	-----

	all senses	main senses only
polysemy:	3	2

	fine-grained	coarse-grained
<u>entropy</u> :	0.694	0.133

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.981 (0.981)	0.995 (0.995)	0.995 (0.995)
best system	0.898 (0.898)	0.978 (0.978)	0.978 (0.978)
average of systems	0.693 (0.630)	0.769 (0.699)	0.769 (0.699)

worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.887 (0.887)	0.978 (0.978)	0.978 (0.978)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.862 (0.353)	0.962 (0.395)	0.962 (0.395)
average of systems	0.607 (0.487)	0.679 (0.545)	0.679 (0.545)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.876 (0.876)	0.978 (0.978)	0.978 (0.978)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.898 (0.898)	0.978 (0.978)	0.978 (0.978)
average of systems	0.735 (0.701)	0.813 (0.776)	0.813 (0.776)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.887 (0.887)	0.978 (0.978)	0.978 (0.978)

Detailed results**seize-v**

number of items in task:	259
--------------------------	-----

	all senses	main senses only
polysemy:	11	9

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	2.806	2.576

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.921 (0.921)	0.921 (0.921)	0.929 (0.929)
best system	0.709 (0.641)	0.709 (0.641)	0.752 (0.680)
average of systems	0.485 (0.440)	0.485 (0.440)	0.520 (0.471)
worst system	0.123 (0.058)	0.123 (0.058)	0.197 (0.093)
best <u>baseline</u>	0.680 (0.680)	0.680 (0.680)	0.710 (0.710)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.353 (0.129)	0.353 (0.129)	0.370 (0.135)
average of systems	0.255 (0.192)	0.255 (0.192)	0.280 (0.207)
worst system	0.123 (0.058)	0.123 (0.058)	0.197 (0.093)
best <u>baseline</u>	0.498 (0.498)	0.498 (0.498)	0.498 (0.498)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.709 (0.641)	0.709 (0.641)	0.752 (0.680)
average of systems	0.599 (0.564)	0.599 (0.564)	0.639 (0.603)
worst system	0.402 (0.402)	0.402 (0.402)	0.483 (0.483)

best <u>baseline</u>	0.680 (0.680)	0.680 (0.680)	0.710 (0.710)
----------------------	----------------------	----------------------	----------------------

Detailed results**brilliant-a**

number of items in task:	229
--------------------------	-----

	all senses	main senses only
polysemy:	10	8

	fine-grained	coarse-grained
entropy:	2.382	1.965

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.930 (0.926)	0.954 (0.950)	0.954 (0.950)
best system	0.563 (0.563)	0.642 (0.642)	0.642 (0.642)
average of systems	0.404 (0.370)	0.474 (0.433)	0.474 (0.433)
worst system	0.062 (0.035)	0.108 (0.061)	0.108 (0.061)
best <u>baseline</u>	0.512 (0.510)	0.610 (0.608)	0.610 (0.608)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.524 (0.485)	0.599 (0.555)	0.599 (0.555)
average of systems	0.256 (0.224)	0.307 (0.267)	0.307 (0.267)
worst system	0.062 (0.035)	0.108 (0.061)	0.108 (0.061)
best <u>baseline</u>	0.476 (0.476)	0.585 (0.585)	0.585 (0.585)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.563 (0.563)	0.642 (0.642)	0.642 (0.642)
average of systems	0.478 (0.443)	0.557 (0.515)	0.557 (0.515)
worst system	0.388 (0.384)	0.427 (0.424)	0.427 (0.424)
best <u>baseline</u>	0.512 (0.510)	0.610 (0.608)	0.610 (0.608)

Detailed results**deaf-a**

number of items in task:	122
--------------------------	-----

	all senses	main senses only
polysemy:	5	5

	fine-grained	coarse-grained
entropy:	1.224	1.224

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.985 (0.985)	0.985 (0.985)	0.985 (0.985)
best system	0.967 (0.967)	0.967 (0.967)	0.967 (0.967)
average of systems	0.808 (0.699)	0.838 (0.727)	0.838 (0.727)
worst system	0.617 (0.331)	0.617 (0.331)	0.617 (0.331)
best <u>baseline</u>	0.902 (0.902)	0.902 (0.902)	0.902 (0.902)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.892 (0.877)	0.892 (0.877)	0.892 (0.877)
average of systems	0.816 (0.646)	0.862 (0.687)	0.862 (0.687)
worst system	0.639 (0.566)	0.824 (0.730)	0.824 (0.730)
best <u>baseline</u>	0.902 (0.902)	0.902 (0.902)	0.902 (0.902)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.967 (0.967)	0.967 (0.967)	0.967 (0.967)
average of systems	0.801 (0.742)	0.819 (0.760)	0.819 (0.760)
worst system	0.617 (0.331)	0.617 (0.331)	0.617 (0.331)
best <u>baseline</u>	0.704 (0.693)	0.704 (0.693)	0.704 (0.693)

Detailed results

floating-a

number of items in task:	47
--------------------------	----

	all senses	main senses only
polysemy:	5	4

	fine-grained	coarse-grained
entropy:	1.748	1.539

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.979 (0.979)	0.979 (0.979)	0.979 (0.979)
best system	0.809 (0.809)	0.809 (0.809)	0.809 (0.809)
average of systems	0.389 (0.373)	0.405 (0.387)	0.408 (0.389)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.660 (0.660)	0.681 (0.681)	0.681 (0.681)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.578 (0.553)	0.578 (0.553)	0.578 (0.553)
average of systems	0.216 (0.197)	0.237 (0.214)	0.241 (0.217)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.596 (0.596)	0.596 (0.596)	0.596 (0.596)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.809 (0.809)	0.809 (0.809)	0.809 (0.809)
average of systems	0.496 (0.484)	0.510 (0.495)	0.512 (0.497)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.660 (0.660)	0.681 (0.681)	0.681 (0.681)

Detailed results**generous-a**

number of items in task:	227
--------------------------	-----

	all senses	main senses only
polysemy:	6	6

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy</u> :	2.303	2.303

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.969 (0.969)	0.969 (0.969)	0.969 (0.969)
best system	0.612 (0.612)	0.612 (0.612)	0.612 (0.612)
average of systems	0.483 (0.434)	0.483 (0.434)	0.483 (0.434)

worst system	0.294 (0.066)	0.294 (0.066)	0.294 (0.066)
best <u>baseline</u>	0.488 (0.488)	0.488 (0.488)	0.488 (0.488)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.464 (0.432)	0.464 (0.432)	0.464 (0.432)
average of systems	0.410 (0.307)	0.410 (0.307)	0.410 (0.307)
worst system	0.294 (0.066)	0.294 (0.066)	0.294 (0.066)
best <u>baseline</u>	0.407 (0.405)	0.407 (0.405)	0.407 (0.405)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.612 (0.612)	0.612 (0.612)	0.612 (0.612)
average of systems	0.519 (0.498)	0.519 (0.498)	0.519 (0.498)
worst system	0.383 (0.383)	0.383 (0.383)	0.383 (0.383)
best <u>baseline</u>	0.488 (0.488)	0.488 (0.488)	0.488 (0.488)

Detailed results**giant-a**

number of items in task:	97
--------------------------	----

	all senses	main senses only
polysemy:	5	2

	<u>fine-grained</u>	<u>coarse-grained</u>
<u>entropy:</u>	0.617	0.214

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
best system	0.990 (0.990)	0.995 (0.995)	1.000 (1.000)
average of systems	0.643 (0.602)	0.685 (0.644)	0.688 (0.647)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.990 (0.990)	0.995 (0.995)	1.000 (1.000)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.571 (0.536)	0.649 (0.649)	0.649 (0.649)
average of systems	0.195 (0.168)	0.301 (0.273)	0.301 (0.273)
worst system	0.000 (0.000)	0.011 (0.010)	0.011 (0.010)
best <u>baseline</u>	0.985 (0.985)	0.995 (0.995)	1.000 (1.000)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.990 (0.990)	0.995 (0.995)	1.000 (1.000)
average of systems	0.867 (0.819)	0.877 (0.830)	0.881 (0.834)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

best <u>baseline</u>	0.990 (0.990)	0.995 (0.995)	1.000 (1.000)
----------------------	----------------------	----------------------	----------------------

Detailed results**modest-a**

number of items in task:	270
--------------------------	-----

	all senses	main senses only
polysemy:	9	3

	fine-grained	coarse-grained
entropy:	2.298	1.323

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.920 (0.920)	0.935 (0.935)	0.941 (0.941)
best system	0.719 (0.719)	0.728 (0.728)	0.768 (0.477)
average of systems	0.566 (0.529)	0.608 (0.569)	0.630 (0.588)
worst system	0.086 (0.085)	0.337 (0.309)	0.368 (0.337)
best <u>baseline</u>	0.648 (0.648)	0.656 (0.656)	0.670 (0.670)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.664 (0.644)	0.669 (0.649)	0.679 (0.659)
average of systems	0.372 (0.315)	0.449 (0.388)	0.476 (0.411)
worst system	0.086 (0.085)	0.337 (0.309)	0.368 (0.337)
best <u>baseline</u>	0.637 (0.637)	0.643 (0.643)	0.656 (0.656)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.719 (0.719)	0.728 (0.728)	0.768 (0.477)
average of systems	0.664 (0.637)	0.687 (0.659)	0.706 (0.676)
worst system	0.567 (0.567)	0.579 (0.579)	0.596 (0.596)
best <u>baseline</u>	0.648 (0.648)	0.656 (0.656)	0.670 (0.670)

Detailed results**slight-a**

number of items in task:	218
--------------------------	-----

	all senses	main senses only
polysemy:	6	3

	fine-grained	coarse-grained
entropy:	1.285	0.432

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.995 (0.995)	0.995 (0.995)	0.995 (0.995)
best system	0.963 (0.963)	0.963 (0.963)	0.963 (0.963)
average of systems	0.715 (0.672)	0.782 (0.730)	0.845 (0.784)
worst system	0.028 (0.028)	0.028 (0.028)	0.028 (0.028)
best <u>baseline</u>	0.954 (0.954)	0.954 (0.954)	0.954 (0.954)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.595 (0.546)	0.728 (0.674)	0.845 (0.775)
average of systems	0.468 (0.389)	0.554 (0.459)	0.660 (0.548)
worst system	0.028 (0.028)	0.028 (0.028)	0.028 (0.028)
best <u>baseline</u>	0.954 (0.954)	0.954 (0.954)	0.954 (0.954)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.963 (0.963)	0.963 (0.963)	0.963 (0.963)
average of systems	0.839 (0.814)	0.896 (0.866)	0.937 (0.903)
worst system	0.682 (0.440)	0.818 (0.528)	0.908 (0.908)
best <u>baseline</u>	0.954 (0.954)	0.954 (0.954)	0.954 (0.954)

[Detailed results](#)

wooden-a

number of items in task:	196
--------------------------	-----

	all senses	main senses only
polysemy:	4	4

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	0.365	0.365

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
best system	0.980 (0.980)	0.980 (0.980)	0.980 (0.980)
average of systems	0.915 (0.849)	0.915 (0.849)	0.915 (0.849)
worst system	0.515 (0.515)	0.515 (0.515)	0.515 (0.515)
best <u>baseline</u>	0.964 (0.964)	0.964 (0.964)	0.964 (0.964)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.958 (0.374)	0.958 (0.374)	0.958 (0.374)
average of systems	0.830 (0.705)	0.830 (0.705)	0.830 (0.705)
worst system	0.515 (0.515)	0.515 (0.515)	0.515 (0.515)
best <u>baseline</u>	0.949 (0.949)	0.949 (0.949)	0.949 (0.949)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.980 (0.980)	0.980 (0.980)	0.980 (0.980)
average of systems	0.958 (0.921)	0.958 (0.921)	0.958 (0.921)
worst system	0.938 (0.934)	0.938 (0.934)	0.938 (0.934)
best <u>baseline</u>	0.964 (0.964)	0.964 (0.964)	0.964 (0.964)

Detailed results**band-p**

number of items in task:	302
--------------------------	-----

	all senses	main senses only
polysemy:	29	25

	fine-grained	coarse-grained
entropy:	1.749	1.669

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.990 (0.990)	0.990 (0.990)	0.990 (0.990)
best system	0.904 (0.904)	0.907 (0.907)	0.907 (0.907)
average of systems	0.616 (0.552)	0.617 (0.552)	0.617 (0.552)

worst system	0.116 (0.116)	0.116 (0.116)	0.116 (0.116)
best <u>baseline</u>	0.852 (0.843)	0.852 (0.843)	0.852 (0.843)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.883 (0.828)	0.883 (0.828)	0.883 (0.828)
average of systems	0.294 (0.228)	0.294 (0.228)	0.295 (0.229)
worst system	0.116 (0.116)	0.116 (0.116)	0.116 (0.116)
best <u>baseline</u>	0.250 (0.248)	0.257 (0.255)	0.257 (0.255)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.904 (0.904)	0.907 (0.907)	0.907 (0.907)
average of systems	0.864 (0.822)	0.865 (0.822)	0.865 (0.822)
worst system	0.831 (0.765)	0.831 (0.765)	0.831 (0.765)
best <u>baseline</u>	0.852 (0.843)	0.852 (0.843)	0.852 (0.843)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.388 (0.114)	0.388 (0.114)	0.388 (0.114)
average of systems	0.388 (0.114)	0.388 (0.114)	0.388 (0.114)
worst system	0.388 (0.114)	0.388 (0.114)	0.388 (0.114)

Detailed results

bitter-p

number of items in task:	373
--------------------------	-----

	all senses	main senses only
polysemy:	14	10

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	2.666	2.472

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.924 (0.917)	0.927 (0.920)	0.927 (0.920)
best system	1.000 (0.008)	1.000 (0.008)	1.000 (0.008)
average of systems	0.563 (0.456)	0.578 (0.468)	0.580 (0.470)
worst system	0.240 (0.233)	0.269 (0.261)	0.273 (0.265)
best <u>baseline</u>	0.551 (0.550)	0.556 (0.555)	0.556 (0.555)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.497 (0.440)	0.526 (0.240)	0.534 (0.244)
average of systems	0.406 (0.323)	0.434 (0.345)	0.437 (0.347)
worst system	0.240 (0.233)	0.269 (0.261)	0.273 (0.265)
best <u>baseline</u>	0.403 (0.402)	0.414 (0.412)	0.414 (0.412)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.729 (0.729)	0.735 (0.735)	0.737 (0.737)
average of systems	0.597 (0.567)	0.608 (0.576)	0.610 (0.578)
worst system	0.475 (0.475)	0.477 (0.477)	0.477 (0.477)
best <u>baseline</u>	0.551 (0.550)	0.556 (0.555)	0.556 (0.555)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	1.000 (0.008)	1.000 (0.008)	1.000 (0.008)
average of systems	1.000 (0.008)	1.000 (0.008)	1.000 (0.008)
worst system	1.000 (0.008)	1.000 (0.008)	1.000 (0.008)

Detailed results**hurdle-p**

number of items in task:	323
--------------------------	-----

	all senses	main senses only
polysemy:	11	8

	fine-grained	coarse-grained
<u>entropy</u> :	2.437	2.019

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.985 (0.985)	0.987 (0.987)	0.987 (0.987)
best system	0.793 (0.793)	0.873 (0.873)	0.873 (0.873)
average of systems	0.405 (0.263)	0.458 (0.308)	0.458 (0.308)
worst system	0.097 (0.093)	0.269 (0.269)	0.269 (0.269)
best <u>baseline</u>	0.334 (0.334)	0.467 (0.467)	0.467 (0.467)

[A](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.568 (0.065)	0.568 (0.065)	0.568 (0.065)
average of systems	0.363 (0.162)	0.428 (0.220)	0.428 (0.220)
worst system	0.097 (0.093)	0.300 (0.300)	0.300 (0.300)
best <u>baseline</u>	0.334 (0.334)	0.467 (0.467)	0.467 (0.467)

[S](#) systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.793 (0.793)	0.873 (0.873)	0.873 (0.873)
average of systems	0.481 (0.432)	0.520 (0.465)	0.520 (0.465)
worst system	0.251 (0.251)	0.269 (0.269)	0.269 (0.269)
best <u>baseline</u>	0.284 (0.283)	0.328 (0.327)	0.328 (0.327)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.324 (0.125)	0.360 (0.139)	0.360 (0.139)
average of systems	0.324 (0.125)	0.360 (0.139)	0.360 (0.139)
worst system	0.324 (0.125)	0.360 (0.139)	0.360 (0.139)

Detailed results**sanction-p**

number of items in task:	431
--------------------------	-----

	<u>all senses</u>	<u>main senses only</u>
polysemy:	7	6

	<u>fine-grained</u>	<u>coarse-grained</u>
entropy:	1.810	1.722

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.981 (0.981)	0.984 (0.984)	0.984 (0.984)
best system	0.865 (0.865)	0.865 (0.865)	0.865 (0.865)
average of systems	0.548 (0.496)	0.554 (0.502)	0.554 (0.502)
worst system	0.094 (0.030)	0.094 (0.030)	0.094 (0.030)
best <u>baseline</u>	0.781 (0.780)	0.781 (0.780)	0.781 (0.780)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.775 (0.703)	0.775 (0.703)	0.775 (0.703)
average of systems	0.340 (0.268)	0.346 (0.272)	0.346 (0.272)
worst system	0.145 (0.023)	0.145 (0.023)	0.145 (0.023)
best <u>baseline</u>	0.585 (0.585)	0.601 (0.601)	0.601 (0.601)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.865 (0.865)	0.865 (0.865)	0.865 (0.865)
average of systems	0.738 (0.703)	0.746 (0.710)	0.746 (0.710)
worst system	0.592 (0.592)	0.617 (0.617)	0.617 (0.617)
best <u>baseline</u>	0.781 (0.780)	0.781 (0.780)	0.781 (0.780)

O systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.094 (0.030)	0.094 (0.030)	0.094 (0.030)
average of systems	0.094 (0.030)	0.094 (0.030)	0.094 (0.030)
worst system	0.094 (0.030)	0.094 (0.030)	0.094 (0.030)

Detailed results

shake-p

number of items in task:	356
--------------------------	-----

	all senses	main senses only
polysemy:	36	30

	fine-grained	coarse-grained
entropy:	3.696	3.531

All systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
<u>human</u>	0.974 (0.974)	0.977 (0.977)	0.978 (0.978)
best system	0.747 (0.747)	0.767 (0.767)	0.781 (0.781)
average of systems	0.496 (0.456)	0.510 (0.469)	0.517 (0.476)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.632 (0.632)	0.640 (0.640)	0.649 (0.649)

A systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.605 (0.328)	0.621 (0.525)	0.631 (0.534)
average of systems	0.314 (0.254)	0.334 (0.270)	0.338 (0.274)
worst system	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
best <u>baseline</u>	0.583 (0.581)	0.609 (0.607)	0.614 (0.612)

S systems

	<u>fine-grained</u> precision (recall)	<u>mixed-grained</u> precision (recall)	<u>coarse-grained</u> precision (recall)
best system	0.747 (0.747)	0.767 (0.767)	0.781 (0.781)
average of systems	0.673 (0.643)	0.684 (0.655)	0.694 (0.665)
worst system	0.559 (0.559)	0.575 (0.575)	0.587 (0.587)
best <u>baseline</u>	0.632 (0.632)	0.640 (0.640)	0.649 (0.649)

Detailed results